Scientifically Based Research: A Planning Tool for Educators

Brought to you by AEL
aelinfo@ael.org or 800-624-9120

This tool accompanies *An Educator's Guide to Scientifically Based Research*. The Guide appeared as a supplement to the January 2004 issue of District Administration Magazine. Its development was sponsored by:

- **Texas Instruments**
- **American Education Corporation**
- **NetTrekker**
- **Inspiration Software.**

Reprint information:
To order printed copies of the Guide, please contact: Lisa Marie Smith, District Administration Magazine (lsmith@promediagrp.com, 203-663-0103)
SBR bulk-copy pricing: 10-50 copies $2.95/copy; 51-200 copies $2.45/copy; 201-1,000 copies $1.95/copy

Scientifically Based Research: A Planning Tool for Educators

Steps to Evaluating Research Conducted by Others

Educators can use the following checklist to evaluate research evidence. This evidence might be the information that is provided by a publisher or program developer. Or, it might be an article about an educational practice. The more questions that can be answered with "yes," the more likely it is that the evidence is scientifically based.

- **Relevance**

- *Does the evidence provided by the researchers or developers address a question that is important to your needs?* For example, if you have disaggregated your student achievement data and it is clear that many fifth-grade students in Title I schools are performing poorly in algebraic concepts, does the evidence provided demonstrate that the product or program under consideration can improve the performance of such students?

- *Do the developers provide evidence that the research that they claim supports their product or program links to and flows from relevant theory and theory-based research?* While you may not have the time or inclination to validate this kind of "linkage," developers should provide evidence that they have documented such linkage. One way they may do this is by conducting a review of existing scientific research related to their product/program. They may also provide a "white paper" that shows the relationship between the literature review and their product/program.

- *Do the research procedures, analyses, and findings support the researchers'/developers' claims?* This can be determined by reviewing the research evidence provided by the developers, checking the U.S. Department of Education's What Works Clearinghouse Web site (www.w-w-c.org), and/or seeking the assistance of research professionals. In some cases, school districts may employ a research staff; in other cases, they may need to draw upon a research firm or university experts.

- **Rigor**

- *If the researchers or developers claim a causal relationship between the intervention (product, service, program) and an outcome measure such as student achievement, did they include a control or comparison group in the study, in addition to the experimental group?*

- *Were the study participants (usually students or teachers or schools) randomly selected and/or randomly assigned to experimental versus control/comparison groups?*

- *Is sufficient information provided to determine whether the research design, instruments, and procedures are appropriate for answering the research questions posed by the researchers/developers?* For example, if the researchers/developers claim that a

particular program improves students' engagement in learning, did they adequately define engagement? Did they provide information about the reliability and validity of the instruments or processes used to measure student engagement? If the researchers/developers claim that a program is effective, did they conduct an experiment or quasi-experiment? Or, did they conduct a survey only? Surveys by themselves do not prove anything. They provide information about what the respondents think or perceive or report.

- *Were the research instruments and procedures applied with consistency, accuracy, and for the purpose intended by the developers of the instruments and procedures?* Researchers should provide enough information for the reader/reviewer to make this judgment. If they do not, then evidence is lacking. Just as research designs should match the purpose of the research study, the instruments used in a research study should be used as they were intended. For example, norm-referenced achievement tests were not originally designed to show how well students measure up against state achievement standards. So, if developers/researchers want to claim that a particular program improves students' performance on the state's standards, then an instrument that was specifically designed to measure achievement of those standards should be used.

- **Systematic Approach**

- *Was the research conducted using carefully planned, logical steps?* Were the steps such that following them logically could lead to answering the research question(s)?

- **Objectivity**

- *Did someone other than the publisher or developer conduct the research attesting to the products or programs effectiveness?* If not, was the research conducted by the publisher/developer submitted to review by an independent, expert panel?

- **Replicability**

- *With the information provided, could the same researchers likely repeat the study and obtain the same or highly similar results?*

- *With the information provided, could other researchers likely replicate the study's methodology and obtain the same or highly similar results?*

- **Data Analyses and Interpretation**

- *Does the research evidence provided include data or data summaries?*

- *Are significance levels and effect sizes reported?* In education, statistically significant findings are generally .05 or less. A significance level indicates the probability that a particular finding is due to chance rather than to the experimental intervention, for example. If the difference between test scores for the experimental group and the control

group is statistically significant at the .05 level, it means there is a five percent probability or "chance" that the findings are erroneous. More important than statistical significance alone, however, are effect sizes. Effect sizes are reported in terms of standard deviation units and tell us something about the practical significance of research findings, i.e., effect sizes are indicators of the size or magnitude of the statistically significant difference between the experimental treatment and control groups. Effect sizes of 1.0 or greater are generally considered large. Effect sizes of .50 are considered "medium," and effect sizes of .25 are considered small, i.e., of little practical significance.

- *Are the conclusions drawn by the researchers/developers clearly supported by the data?* If no data or data summaries, significance levels, or effect sizes are provided, it will be difficult, if not impossible, to answer this question.

Steps to Conducting Scientifically Based Research

In some cases, school districts may be interested in developing an educational program of their own. If the program is dependent upon a funding source that requires scientifically based research (SBR), then districts will need to follow certain steps. The following list describes the steps to conducting SBR aimed at demonstrating cause-and-effect relationships. The emphasis is on causation because it is the primary and immediate concern of schools, districts, and states striving to comply with NCLB requirements. This guidance can be used whether you will be conducting your own research or contracting for research services. It can also be used to assess the extent to which companies developing commercial products employed the best research practices.

1. *Formulate a hypothesis* about the effect of the independent or "causal" variable (such as a particular instructional strategy) on the dependent or outcome variable (such as student achievement). This hypothesis should be based upon the best available information (e.g., sound theory, prior rigorous research, and/or empirical observation). A sample hypothesis might be: When third-grade students are exposed to 100 hours of XYZ software for increasing reading comprehension, their scores on a test of reading comprehension will increase.

2. *Randomly select a sample* of participants for the study, if possible. In other words, select participants by using a table of random numbers or by drawing their names "out of a hat," instead of allowing them to volunteer. Also, if possible, randomly assign individual members from the sample to either the experimental or the control/comparison group(s). NCLB places particular emphasis on random assignment. If random selection and/or assignment are possible, you will have the makings of an experimental study. If not, then you will be conducting a quasi-experiment. Either way, you must have both an experimental group and a control or comparison group.

3. *Administer a pretest* to both the experimental and control/comparison groups if you are interested in measuring change over time. This is especially important if you are unable to randomly assign participants to groups. Be sure that the pretest is reliable and valid for the purpose at hand. Information about the reliability and validity of commercially available instruments should be available in the technical manual accompanying such instruments or in reference books such as *Buros Mental Measurements Yearbook* (Plake & Impara (2001) or *Tests in Print* (Murphy et al., 2002). If you are developing your own instruments, someone with expertise and experience in instrument development will need to conduct studies to establish the reliability and validity of these instruments.

4. *Apply the treatment intervention* to the experimental group, being careful to plan and document the nature, specific elements, length, intensity, and context of the treatment. This will allow for replication.

5. *Re-measure* (i.e., "posttest") both the experimental and control/comparison groups, using the pretest measure or a measure that has been demonstrated statistically to be equivalent to the pretest measure. It is important to know or document the reliability of the measures. If the same measure is used for pre and posttesting, then "test-retest reliability" is important. If different measures are used, then "parallel" or "alternate forms" reliability is important. In either case, if the appropriate type of reliability is not reported by the test publisher and you do not have a research staff, researchers experienced in instrument development can help you establish the appropriate reliabilities.

6. *Analyze the results* of the measurements of the experimental and control/comparison groups on the pre- and posttest measures. A statistics specialist can help determine the most appropriate types of statistical analyses and tests to conduct. Ultimately, significance levels and effect sizes should be calculated. Effect sizes indicate the practical significance of statistical findings. Large effect sizes tend to be 1.0 or greater. Effect sizes of .50 or so are considered medium, and effect sizes of .25 or less are generally considered small.

7. *Write a report* of the findings that includes a description of (1) the rationale for the study; (2) findings from prior research that contributed to the study's underlying hypothesis; (3) the research procedures and instruments that were used, including information about their reliability and validity; (4) demographic information about the participants in the study, as well as information about how they were selected and how they were assigned to groups; (5) how the results were analyzed; (6) the results of the analyses, including effect sizes; and (7) conclusions that can be supported by the data yielded by the study.

About AEL

Founded in 1966 as a not-for-profit corporation, AEL provides rigorous research, professional development, and consulting services to clients in the education arena. Services include research design and implementation,

comprehensive research reviews, intensive product and program evaluations and field trials, consulting services, and award-winning professional development programs.

The name AEL originated with the first program the corporation operated — the Appalachia Educational Laboratory — but today's AEL is national in scope and provides a range of services to private and government agencies. For information about AEL, visit the AEL Web site (www.ael.org) or contact AEL at aelinfo@ael.org or 800-624-9120.

About the Author

Dr. Doris Redfield is Vice President for Research and Director of the Regional Educational Laboratory at AEL, one of 10 federally funded educational research and development laboratories. AEL works with education practitioners, researchers, and policymakers to study and promote strategies for student success.